Epidemic dynamics in finite size scale-free networks

Romualdo Pastor-Satorras[1] and Alessandro Vespignani[2]

[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*
[2]*The Abdus Salam International Centre for Theoretical Physics (ICTP), P.O. Box 586, 34100 Trieste, Italy*
(Dated: February 1, 2008)

Many real networks present a bounded scale-free behavior with a connectivity cut-off due to physical constraints or a finite network size. We study epidemic dynamics in bounded scale-free networks with soft and hard connectivity cut-offs. The finite size effects introduced by the cut-off induce an epidemic threshold that approaches zero at increasing sizes. The induced epidemic threshold is very small even at a relatively small cut-off, showing that the neglection of connectivity fluctuations in bounded scale-free networks leads to a strong over-estimation of the epidemic threshold. We provide the expression for the infection prevalence and discuss its finite size corrections. The present work shows that the highly heterogeneous nature of scale-free networks does not allow the use of homogeneous approximations even for systems of a relatively small number of nodes.

In the last years it has been recognized that a large number of physical, biological, and social networks exhibits complex topological properties [1, 2]. In particular many real world networks show the small-world phenomenon, related to a very small average path length between nodes [2, 3]. More strikingly, in some cases this property is associated to a scale-free connectivity distribution, $P(k) \sim k^{-2-\gamma}$, with $0 < \gamma \leq 1$, where k is the number of links connected to a node [4]. This scale-free nature is associated to a large heterogeneity in the connectivity properties of the system. Since the second moment of the connectivity distribution $\langle k^2 \rangle$ is diverging when increasing the network size, the connectivity fluctuations in scale-free (SF) networks do not have an intrinsic bound and diverge in the infinite system size limit. Scale-free properties have been observed in many real systems such as the Internet [5, 6, 7] and the World-Wide-Web [4, 8], food-webs, protein, and neural networks [9]. A very important example of scale-free networks is also found in the web of human sexual contacts [10]. This is a particularly relevant case since the unambiguous definition of contacts (links) is often missing in the analysis of social networks.

Since the Internet and the web of human sexual contacts appear to be scale-free, the study of epidemics and disease dynamics on SF networks is a relevant theoretical issue in the spreading of computer viruses and sexually transmittable diseases. In heterogeneous networks, it is well-known that the epidemic threshold decreases with the standard deviation of the connectivity distribution [11, 12], and this feature is amplified in SF networks, which have diverging connectivity fluctuations in the limit of infinite network size. Indeed, it was first noted in Ref. [13] that, in infinite SF networks, epidemic processes do not possess an epidemic threshold below which diseases cannot produce a major epidemic outbreak or the inset of an endemic state. The absence of an intrinsic epidemic threshold has been found in both the susceptible-infected-susceptible (SIS) model [13] and the susceptible-infected-removed (SIR) model [14, 15] in infinite SF networks. The immunization policies are as well very much affected by the SF nature of the connectivity distribution [16, 17].

As customarily encountered in nonequilibrium statistical systems [18], it has also been pointed out that in finite systems an epidemic threshold is induced by finite size effects [14]. Real systems are actually made up by a finite number of individuals which is far from the thermodynamic limit. This finite population introduces a maximum connectivity k_c, depending on N, which has the effect of restoring a bound in the connectivity fluctuations, inducing in this way an effective nonzero threshold. More generally, we can consider a class of bounded scale-free (BSF) networks, in which the connectivity distribution has the form $P(k) \sim k^{-2-\gamma} f(k/k_c)$, where the function $f(x)$ decreases very rapidly for $x > 1$ [19]. The cut-off k_c can be due to the finite size of the network or to the presence of constraints limiting the addition of new links in an otherwise infinite network [1].

In this paper we present an analytical study of the SIS model in BSF networks with a generic connectivity exponent γ ($0 < \gamma \leq 1$), focusing on the effects introduced by a finite cut-off k_c. We analyze the case of a hard cut-off, $f(x) = \theta(1-x)$, where $\theta(x)$ is the Heaviside step function, as it happens in growing networks with a finite number of elements. We consider as well a soft exponential cut-off, $f(x) = \exp(-x)$, as often found in systems where physical constraints are at play. We derive the behavior of the epidemic threshold as a function of k_c and the network size N, and find that even for relatively small networks the induced epidemic threshold is much smaller than the epidemic threshold found in homogeneous systems. This confirms that the SF nature cannot be neglected in the practical estimates of epidemic and immunization thresholds in real networks. We also provide the explicit analytic form for the epidemic prevalence (density of infected individuals) in BSF networks. The results presented here can be readily extended to the SIR case.

In order to estimate the effect of k_c in epidemics

on BSF networks we will investigate the standard SIS model [20]. This model relies on a coarse-grained description of individuals in the population. Namely, each node of the graph represents an individual and each link is a connection along which the infection can spread. Each susceptible (healthy) node is infected with rate ν if it is connected to one or more infected nodes. Infected nodes are cured and become again susceptible with rate δ, defining an effective spreading rate $\lambda = \nu/\delta$ (without lack of generality, we set $\delta = 1$). The SIS model does not take into account the possibility of individuals removal due to death or acquired immunization [20], and individuals run stochastically through the cycle susceptible $\rightarrow$ infected $\rightarrow$ susceptible. This model is generally used to study infections leading to endemic states with a stationary average density of infected individuals. In order to take into account the heterogeneity of SF networks, we have to relax the homogeneity assumption used in regular networks, and consider the relative density $\rho_k(t)$ of infected nodes with given connectivity k; i.e., the probability that a node with k links is infected [13]. The dynamical mean-field equations can thus be written as

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda k \left[1 - \rho_k(t)\right]\Theta(\rho(t)). \qquad (1)$$

The first term in Eq. (1) considers infected nodes becoming healthy with unit rate. The second term represents the average density of newly generated infected nodes that is proportional to the infection spreading rate λ and the probability that a node with k links is healthy $[1 - \rho_k(t)]$ and gets the infection via a connected node. The rate of this last event is given by the probability $\Theta(\rho(t))$ that any given link points to an infected node, which has the expression [13]

$$\Theta(\rho(t)) = \langle k \rangle^{-1} \sum_k k P(k) \rho_k(t). \qquad (2)$$

By solving Eqs. (1) and (2) in the stationary state $[d\rho_k(t)/dt = 0]$ we obtain the self-consistency equation [13]

$$\Theta = \langle k \rangle^{-1} \sum_k k P(k) \frac{\lambda k \Theta}{1 + \lambda k \Theta}, \qquad (3)$$

where Θ is now a function of λ alone. The self-consistency Eq. (3) allows a solution with $\Theta \neq 0$ and $\rho_k \neq 0$ only if the condition $\lambda \langle k^2 \rangle / \langle k \rangle \geq 1$ is fulfilled [16], defining the epidemic threshold

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}. \qquad (4)$$

In other words, if the value of λ is above the threshold, $\lambda \geq \lambda_c$, the infection spreads and becomes endemic. Below it, $\lambda < \lambda_c$, the infection dies out exponentially fast. This result implies that in infinite SF networks with connectivity exponent $0 < \gamma \leq 1$, for which $\langle k^2 \rangle \rightarrow \infty$, we have $\lambda_c = 0$. This fact implies in turn that for any positive value of λ the infection can pervade the system with a finite prevalence, in a sufficiently large network [13]. While this results is valid for infinite SF networks, $\langle k^2 \rangle$ assumes a finite value in BSF networks, defining an effective nonzero threshold due to finite size effects as usually encountered in nonequilibrium phase transitions [18]. This epidemic threshold, however, is not an *intrinsic* quantity as in homogeneous systems and it vanishes for a increasing network size or connectivity cut-off. In order to calculate the precise effects of a finite k_c we consider two different cases of connectivity cut-off. At first instance we consider a *soft* exponential cut-off with *characteristic* connectivity k_c. This case corresponds to those real networks in which external factors set up an upper limit to the connectivity [1]. The network can have an infinite number of elements but the power-law connectivity distribution decays exponentially for large values of k. In order to perform explicit calculations we use a continuous approximation that substitutes the connectivity by a real variable k in the range $[m, \infty)$, where m is the minimum connectivity of the network. The connectivity probability distribution in this case is $P(k) = Ak^{-2-\gamma} \exp(-k/k_c)$, where A is a normalization factor. The effective nonzero epidemic threshold $\lambda_c(k_c)$ induced by the exponential cut-off is given by

$$\lambda_c(k_c) = \frac{\int_m^\infty k^{-1-\gamma} \exp(-k/k_c) dk}{\int_m^\infty k^{-\gamma} \exp(-k/k_c) dk}, \qquad (5)$$

which, after integration, yields

$$\lambda_c(k_c) = k_c^{-1} \frac{\Gamma(-\gamma, m/k_c)}{\Gamma(1-\gamma, m/k_c)}, \qquad (6)$$

where $\Gamma(x, y)$ is the incomplete Gamma function [21]. For large k_c we can perform a Taylor expansion and retain only the leading term, obtaining for any $0 < \gamma < 1$

$$\lambda_c(k_c) \simeq \frac{1}{m\gamma\Gamma(1-\gamma)}(k_c/m)^{\gamma-1}. \qquad (7)$$

The limit $\gamma \rightarrow 1$ in Eq. (6) corresponds to a logarithmic divergence, yielding at leading order $\lambda_c(k_c) \simeq (m\ln(k_c/m))^{-1}$. In all cases we have that the epidemic threshold vanishes when increasing the characteristic cut-off. For large k_c, the average connectivity is virtually fixed and given by $\langle k \rangle = (\gamma + 1)m/\gamma$, for any $\gamma > 0$. It is interesting thus to compare the intrinsic epidemic threshold obtained in homogeneous networks with negligible fluctuations and the nonzero effective threshold of BSF networks. The intrinsic epidemic threshold of homogeneous networks with constant node connectivity $\langle k \rangle$ is given by $\lambda_c^{\mathrm{H}} = \langle k \rangle^{-1}$ [13, 20]. If we compare BSF and homogeneous networks with the same average connectivity $\langle k \rangle = (\gamma + 1)m/\gamma$ we obtain that the ratio between the epidemic thresholds is given by

$$\frac{\lambda_c(k_c)}{\lambda_c^{\mathrm{H}}} \simeq \frac{(\gamma + 1)}{\gamma^2 \Gamma(1-\gamma)}(k_c/m)^{\gamma-1}. \qquad (8)$$
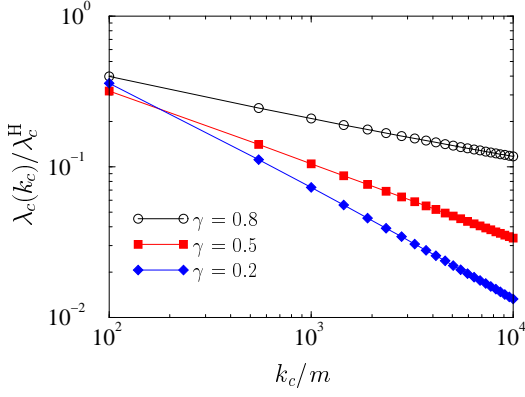
FIG. 1: Ratio between the effective epidemic threshold in BSF networks with a soft exponential cut-off k_c and the intrinsic epidemic threshold of homogeneous networks with the same average connectivity, for different values of γ.

This clearly shows that even in the case of a connectivity cut-off the effective epidemic threshold in BSF networks is much smaller than the intrinsic threshold obtained in regular networks. In Fig. 1 we plot the ratio obtained by using the full expression for $\lambda_c(k_c)$, Eq. (6). It is striking to observe that, even with relatively small cut-offs ($k_c \sim 10^2 - 10^3$), for $\gamma \approx 0.5$ the effective epidemic threshold of BSF networks is smaller by a factor close to $1/10$ than the intrinsic threshold obtained on homogeneous networks.

As a second kind of finite size effect we consider the presence of a *hard* cut-off k_c. Since SF networks are often dynamically growing networks, this case represent a network which has grown up to a finite number of nodes N. The maximum connectivity k_c of any node is related to the network age, measured as the number of nodes N, by the scaling relation [19]

$$k_c \simeq mN^{1/(1+\gamma)}, \tag{9}$$

where m is the minimum connectivity of the network.

In this case the network does not possess any node with connectivity k larger than k_c, and we can think in terms of a hard cut-off. Using again the continuous k approximation, the normalized connectivity distribution has now the form

$$P(k) = \frac{(1+\gamma)m^{1+\gamma}}{1 - (k_c/m)^{-1-\gamma}} k^{-2-\gamma}\theta(k_c - k), \tag{10}$$

where $\theta(x)$ is the Heaviside step function. The finite size induced epidemic threshold $\lambda_c(k_c)$ is given by the expression

$$\lambda_c(k_c) = \frac{\int_m^{k_c} k^{-1-\gamma}dk}{\int_m^{k_c} k^{-\gamma}dk}. \tag{11}$$

Evaluating the above expression we obtain at leading order in k_c/m:

$$\lambda_c(k_c) \simeq \frac{1-\gamma}{\gamma m}(k_c/m)^{\gamma-1}. \tag{12}$$

In this case the hard cut-off k_c can be expressed as a function of the network size N by using the scaling relation Eq. (9) and we can obtain the effective epidemic threshold as

$$\lambda_c(N) \simeq \frac{1-\gamma}{\gamma m}N^{(\gamma-1)/(\gamma+1)}. \tag{13}$$

This expression is valid for any $0 < \gamma < 1$, while for $\gamma = 1$ we obtain at the leading order the logarithmic behavior $\lambda_c(N) \simeq 2(m\ln(N))^{-1}$. Also in this case we have that the effective epidemic threshold is approaching zero for increasing network sizes, and it is worth comparing its magnitude with the corresponding intrinsic threshold in homogeneous networks with identical average connectivity. In Fig. 2 we report the ratio $\lambda_c(N)/\lambda_c^H$ for different sizes of the SF network. It is striking to notice that for $\gamma = 0.5$, small networks with $N \simeq 10^4$ exhibit a finite size induced epidemic threshold that is close to be one order of magnitude smaller than the intrinsic epidemic threshold of a homogeneous network.

In order to find the prevalence behavior we have to solve Eq. (3) in the continuous approximation,

$$\Theta = \langle k \rangle^{-1} \int_m^\infty kP(k)\frac{\lambda\Theta k}{1+\lambda\Theta k}dk, \tag{14}$$

and use the value of Θ to compute the density of infected sites ρ as

$$\rho = \sum_k \rho_k P(k) \equiv \int_m^\infty P(k)\frac{\lambda\Theta k}{1+\lambda\Theta k}dk, \tag{15}$$

where $P(k)$ is given by Eq. (10). In the absence of any cut-off ($k_c \to \infty$) and in the thermodynamic limit ($N \to \infty$) the prevalence scales as $\rho \sim \lambda^{1/(1-\gamma)}$ if $0 < \gamma < 1$, and as $\rho \sim \exp(-1/m\lambda)$ if $\gamma = 1$ [13]. Accordingly with the absence of the epidemic threshold, the prevalence is null only if the spreading rate is $\lambda = 0$. In the case of a hard cut-off we can integrate Eq. (14), neglecting terms of order $(k_c/m)^{-\gamma}$ in the $P(k)$ distribution, to obtain:

$$\begin{aligned}
\Theta &\simeq \gamma m^\gamma \lambda\Theta \int_m^{k_c} \frac{k^{-\gamma}}{1+\lambda\Theta k}dk \\
&= F(1,\gamma,1+\gamma,-[\lambda\Theta m]^{-1}) \\
&\quad -(k_c/m)^{-\gamma}F(1,\gamma,1+\gamma,-[\lambda\Theta k_c]^{-1}),
\end{aligned}$$

where F is the Gauss hyper-geometric function [21]. For a *fixed* k_c one can expand both hyper-geometric functions on the right hand side in the previous equation, keeping the most relevant terms in Θ and considering afterwards the limit of large k_c. The final solution for Θ is then given, at leading order in (k_c/m), by

$$\Theta \simeq \frac{1}{m\lambda^2}\frac{2-\gamma}{1-\gamma}\left(\frac{k_c}{m}\right)^{-1}\left[\lambda - \frac{1-\gamma}{\gamma m}\left(\frac{k_c}{m}\right)^{\gamma-1}\right]. \tag{16}$$

By evaluating the integral in Eq. (15) and keeping the leading term in Θ and k_c we finally obtain the infection
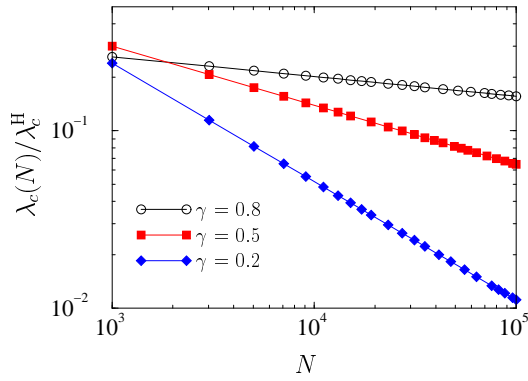
FIG. 2: Ratio between the effective epidemic threshold in BSF networks with finite size N and the intrinsic epidemic threshold of homogeneous networks with the same average connectivity, for differen values of γ.

prevalence as

$$\rho \simeq \frac{(\gamma+1)(2-\gamma)}{\lambda\gamma(1-\gamma)}\left(\frac{k_c}{m}\right)^{-1}\left[\lambda-\frac{1-\gamma}{\gamma m}\left(\frac{k_c}{m}\right)^{\gamma-1}\right].$$

Inserting the scaling relation Eq. (9) between the maximum connectivity k_c and the network size N we are led to the final expression

$$\rho \sim N^{-1/(\gamma+1)}(\lambda-\lambda_c(N)). \qquad (17)$$

That is, the finite size of the network induces a standard mean-field transition at the induced epidemic threshold $\lambda_c(N)$, given by Eq. (13). As can be seen from Eq. (17), however, the prevalence is depressed by a factor $N^{-1/(\gamma+1)}$ from the corresponding value for a homogeneous network. The above calculations can be repeated along similar lines in the case of a *soft* exponential cut-off, obtaining similar results.

It is worth remarking that similar results hold as well for the SIR model. Despite this model confers permanent immunity and does not allow for a stationary state, the epidemic threshold over which an epidemic outbreak occurs has the same analytic form $\lambda_c = \langle k\rangle/\langle k^2\rangle$ [15]. Thus, the present results for the effect of finite size and the induced epidemic threshold can be readily exported to the SIR case. The calculation of the epidemic prevalence is different due to the different evolution equations, but recovers the same onset of an induced mean-field transition at the effective threshold $\lambda_c(N)$. Interestingly, similar results have been obtained in the analysis of the resilience to damage of finite size scale-free networks [22, 23]

In conclusion, we have shown that the SF networks weakness to epidemic agents is also present in finite size networks. Using the homogeneity assumption in the case of SF networks will lead to a serious over-estimate of the epidemic threshold even for relatively small networks.

[1] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. USA **97**, 11149 (2000).
[2] S. H. Strogatz, Nature **410**, 268 (2001).
[3] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[6] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[7] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[8] R. Albert, H. Jeong, and A.-L. Barabási, Nature **401**, 130 (1999).
[9] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[10] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature **411**, 907 (2001).
[11] H. W. Hethcote and J. A. Yorke, Lect. Notes Biomath. **56**, 1 (1984).
[12] R. M. Anderson and R. M. May, *Infectious diseases in humans* (Oxford University Press, Oxford, 1992).
[13] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001). R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).

[14] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[15] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, *Epidemic outbreaks in complex heterogeneous networks* (2001), e-print cond-mat/0107267.
[16] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002).
[17] Z. Dezsö and A.-L. Barabási, *Can we stop the AIDS epidemic?* (2001), e-print cond-mat/0107420.
[18] J. Marro and R. Dickman, *Nonequilibrium phase transitions in lattice models* (Cambridge University Press, Cambridge, 1999).
[19] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks* (2001), e-print cond-mat/0106144.
[20] O. Diekmann and J. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation* (John Wiley & Sons, New York, 2000).
[21] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions.* (Dover, New York, 1972).
[22] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[23] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).